Developmental Science

Developmental Science (2012), pp 1–11 DOI: 10.1111/j.1467-7687.2012.01179.x

PAPER

Neural correlates of perceptual narrowing in cross-species face-voice matching

Tobias Grossmann,[1,2] Manuela Missana,[1] Angela D. Friederici[1] and Asif A. Ghazanfar[3,4,5]

1. *Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
2. *Centre for Brain and Cognitive Development, Birkbeck, University of London, UK*
3. *Neuroscience Institute, Princeton University, USA*
4. *Department of Psychology, Princeton University, USA*
5. *Department of Ecology & Evolutionary Biology, Princeton University, USA*

Abstract

Integrating the multisensory features of talking faces is critical to learning and extracting coherent meaning from social signals. While we know much about the development of these capacities at the behavioral level, we know very little about the underlying neural processes. One prominent behavioral milestone of these capacities is the perceptual narrowing of face–voice matching, whereby young infants match faces and voices across species, but older infants do not. In the present study, we provide neurophysiological evidence for developmental decline in cross-species face–voice matching. We measured event-related brain potentials (ERPs) while 4- and 8-month-old infants watched and listened to congruent and incongruent audio-visual presentations of monkey vocalizations and humans mimicking monkey vocalizations. The ERP results indicated that younger infants distinguished between the congruent and the incongruent faces and voices regardless of species, whereas in older infants, the sensitivity to multisensory congruency was limited to the human face and voice. Furthermore, with development, visual and frontal brain processes and their functional connectivity became more sensitive to the congruence of human faces and voices relative to monkey faces and voices. Our data show the neural correlates of perceptual narrowing in face–voice matching and support the notion that postnatal experience with species identity is associated with neural changes in multisensory processing (Lewkowicz & Ghazanfar, 2009).

Introduction

Human communication in its most natural form is multisensory and multidimensional. For the developing infant, the redundant cues from different sensory modalities can enhance learning and discrimination (Bahrick & Lickliter, 2002). Thus, the infant's brain need not rely solely on an individual's voice or face for learning an identity, but can take advantage of spatial and temporal cues that are common across the two modalities. How are such multisensory signals processed in the developing brain? In cat and monkey model systems, regions of the neocortex and superior colliculus that are multisensory in the adult only gradually acquire this characteristic (Wallace, Perrault, Hairston & Stein, 2004; Wallace & Stein, 1997, 2001; Wallace, Stein & Ramachandran, 2006). In humans, we know very little about multisensory circuit development, and model systems using very precocial animals like macaque monkeys may not accurately represent the developmental processes in the human brain (Zangenehpour, Ghazanfar, Lewkowicz & Zatorre, 2009).

One puzzling feature of human multisensory perceptual development is that it undergoes a period of 'narrowing', whereby young infants respond to native as well as non-native multisensory attributes. As development proceeds, however, infants are selectively exposed to only native perceptual attributes, and this experience reduces ('narrows') the infants' capacity to bind multisensory features in non-native signals (Lewkowicz & Ghazanfar, 2009). For example, 4- and 6-month-old infants can match non-native monkey faces and voices, while 8- and 10-month-old infants exhibit no evidence of multisensory matching (Lewkowicz & Ghazanfar, 2006). These perceptual data suggest that developing circuits in the infant brain are fundamentally multisensory (Ghazanfar, 2012; Hyde, Jones, Flom & Porter, 2011; Hyde, Jones, Porter & Flom, 2010; Lewkowicz & Ghazanfar, 2009).

That infants undergo a period of multisensory perceptual narrowing raises questions regarding its neurodevelopmental correlates. What is the form of audio-visual face/voice neural responses in younger versus older infants? Does congruence matter at either

Address for correspondence: Tobias Grossmann, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany; e-mail: grossman@cbs.mpg.de

age for modulating the magnitude of these responses? Finally, and most importantly for perceptual narrowing, are audio-visual signals of another species processed differently from human signals at different ages? Here, we sought to address these questions by measuring event-related brain potentials in infants at two different ages – 4 months and 8 months – that span the time over which perceptual narrowing for human versus monkey signals occurs. We recorded infants' brain activity while they watched and listened to audio-visual presentations of two monkey vocalizations and two humans mimicking monkey vocalizations (Figure 1). The audio-visual signals were either congruent or incongruent. Incongruent trials consisted of pairing a monkey coo sound with a monkey grunt face, and vice versa, and pairing a human coo sound with a human grunt face, and vice versa. We chose to have humans mimic monkey vocalizations to ascertain that the visual dynamics of the face and acoustic characteristics of the sounds (and thus the vocal tract dynamics) were similar across species. This ensured that any putative differences we might observe were due to species identity, not to low-level stimulus characteristics. Critically, in terms of stimulus selection and age groups tested, our ERP study closely followed a prior behavioral study that produced multisensory perceptual narrowing effects (Lewkowicz & Ghazanfar, 2006).

There were two predictions. First, ERP responses will be differentially modulated by congruent versus incongruent stimuli, perhaps showing greater effects in older infants. Second, and most pertinent to the perceptual narrowing framework, ERP responses should reflect species differences in older but not younger infants. The null hypothesis is that there would be no discernible ERP response changes related to multisensory processes across the two age groups. To be more specific with regard to our electrophysiological predictions, we will first summarize what is currently known about infant ERP correlates of (a) the perceptual processing of human and monkey faces and (b) the processing and integration of combined information from faces and voices.

In adults, human faces elicit an N170 response, which is most prominent over posterior temporal sites and is larger in amplitude to inverted than to upright faces (Bentin, Allison, Puce, Perez & McCarthy, 1996; de Haan, Pascalis & Johnson, 2002). This component is neither modulated by the inversion of monkey faces (de Haan *et al.*, 2002), nor when upright objects are compared to inverted objects (Bentin *et al.*, 1996). This human face-selective effect is evidence that the N170 represents a special face-processing mechanism. From studies examining the influence of face inversion on infants' ERP responses, it has been suggested that the infant N290 response is a precursor to the adult N170. Like the adult N170, the infant N290 is a negative-going deflection observed over posterior electrodes. Its peak latency decreases from 350 ms at 3 months to 290 ms at 12 months of age (Halit *et al.*, 2003).

In the studies that measured infant ERPs to upright and inverted human versus monkey faces (de Haan *et al.*, 2002; Halit *et al.*, 2003), the amplitude of the infant N290 at 12 months of age, like the adult N170, was enhanced to inverted human but not monkey faces when compared to upright faces. However, in line with perceptual narrowing accounts, the amplitude of the N290 was not affected by stimulus inversion at an earlier age (3- and 6-month-olds). These younger infants showed an inversion effect on the amplitude of the *P400* that follows the N290 (Halit *et al.*, 2003). However, unlike the adult N170, the modulation of the amplitude of the P400 at 3 and 6 months of age is not specific to inverted human faces (de Haan *et al.*, 2002; Halit

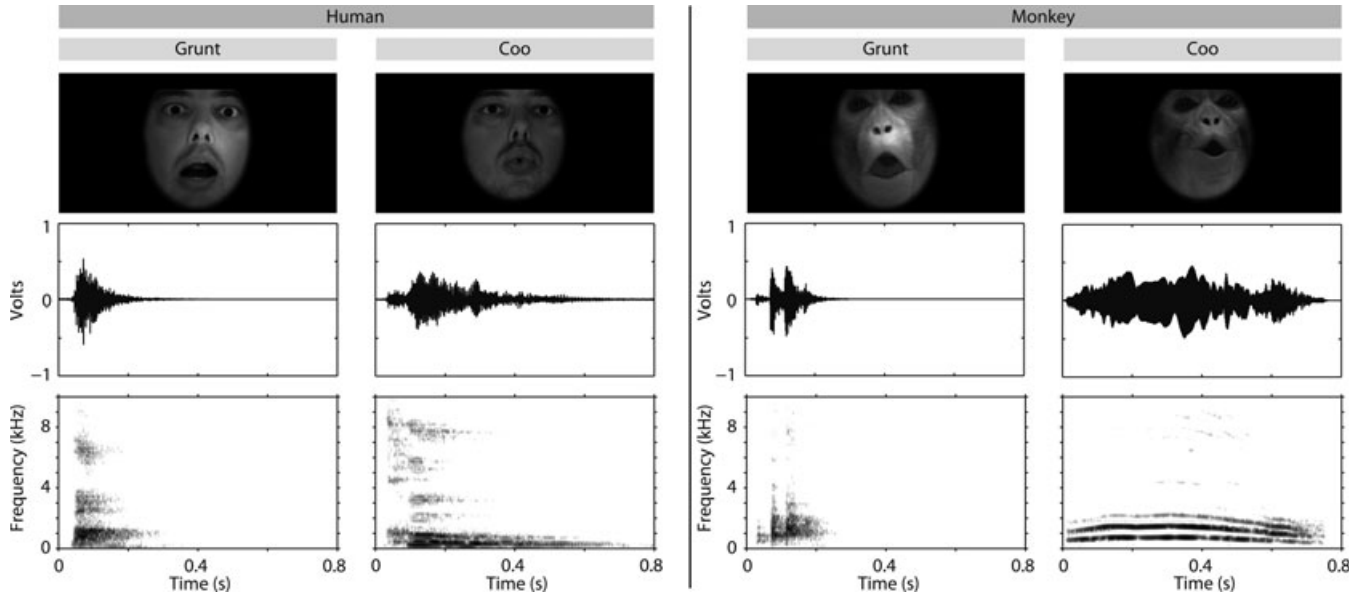

Figure 1 *Representative examples of the stimuli. (Top row) Single video frames of facial gestures made during a coo and a grunt vocalization for the human stimuli (left) that were matched to the monkey stimuli (right), shown at the point of maximal mouth opening. Oscillograms (Middle row) and spectrograms (Middle row) of the vocalizations.*

et al., 2003). It is also important to note that a similar amplitude enhancement for faces compared to matched visual noise was observed in the infant N290 and adult N170, but not for the P400, which makes it unlikely that the P400 is the main precursor of the adult N170 (Halit, Csibra, Volein & Johnson, 2004). Taken together, the prior studies reviewed above suggest that perceptual narrowing to human faces (Pascalis, de Haan & Nelson, 2002) is reliably associated with changes in infants' N290 ERP response. According to these findings, we predict that the visual processing of faces during our multisensory stimulus presentations, as indexed by the infant N290, will become more specialized to human faces as development progresses.

Compared to the extensive research on ERP correlates of face processing during infancy, there is little ERP work on face–voice processing in infancy, and this work is limited to human stimuli and tests infants at an age prior to perceptual narrowing. For example, Kushnerenko and colleagues (2008) measured ERPs in 5-month-olds and demonstrated that brain responses at frontal and temporal electrodes differed (starting around 290 ms) depending on whether conflicting auditory and visual speech cues can or cannot be integrated. Also in 5-month-olds, Hyde and colleagues (2011) showed that asynchronous versus synchronous face–voice pairings elicited a greater (more negative) early visual ERP responses (peak around 150 ms) and a greater negative component at fronto-central electrodes at a later time (peak around 600 ms). These data suggest that there are greater demands during both early sensory and later attentional stages of processing unfamiliar (asynchronous) information across faces and voices. Moreover, at temporal electrodes, synchronous versus asynchronous face–voice pairings elicited greater ERP responses during early (peak around 250 ms) and late stages (after movement offset) of face–voice processing. Interestingly, these synchrony effects were observed over the left hemisphere and might be correlates of brain processes related to processing familiar audio-visual speech information. Based on these prior face–voice studies, we hypothesize that the discrimination of congruent and incongruent face–voice pairs in our study will be reflected in ERP differences during early sensory (occipital and temporal) and later attentional (fronto-central) processing. It is important to emphasize that our study differs in important ways from these prior studies by (a) presenting cross-species stimuli (humans and monkeys) and (b) by testing infants at different ages (4 months and 8 months) that span the time over which perceptual narrowing occurs.

Experimental procedures

Participants

The final sample consisted of 16 4-month-old infants (eight females, M = 126 days, *Range* = 123 to 133 days) and 16 8-month-old infants (seven females, M = 256 days, *Range* = 241 to 268 days). An additional 20 4-month-olds and an additional 23 8-month-olds were tested but not included in the final sample due to too much movement or too many artifacts. The mean number of completed trials for the group of 4-month-olds in the four experimental conditions were: (1) human congruent M = 20.25 (SD = 3.17), (2) human incongruent M = 18.25 (SD = 3.19), (3) monkey congruent M = 19.44 (SD = 3.39), and (4) monkey incongruent M = 18.9 (SD = 3.1). The mean number of completed trials for the group of 8-month-olds were: (1) human congruent M = 21.56 (SD = 2.99), (2) human incongruent M = 19.68 (SD = 3.04), (3) monkey congruent M = 20.81 (SD = 3.08), and (4) monkey incongruent M = 19.5 (SD = 3.03). The mean number of trials included in the ERP average for the group of 4-month-olds were: (1) human congruent M = 11.19 (SD = 2.76), (2) human incongruent M = 11.5 (SD = 2.58), (3) monkey congruent M = 12.7 (SD = 2.59), and (4) monkey incongruent M = 11.9 (SD = 2.19). The mean number of trials included in the ERP average for the group of 8-month-olds were: (1) human congruent M = 11.13 (SD = 3.2), (2) human incongruent M = 11.56 (SD = 2.71), (3) monkey congruent M = 12.31 (SD = 3.68), and (4) monkey incongruent M = 12.68 (SD = 2.47). There were no statistical differences in the number of trials across ages. Note that an attrition rate at this level is within the normal range for an infant ERP study (DeBoer, Scott & Nelson, 2005). All infants were born full-term (37–42 weeks gestation) with normal birthweight (>2500 g). All parents gave informed consent before the study.

Stimuli

The stimuli were digital video recordings of monkeys and humans producing vocalizations. The monkey videos were identical to those used in a behavioral multisensory narrowing study with human infants (Lewkowicz & Ghazanfar, 2006). They consisted of rhesus monkeys (*Macaca mulatta*) producing two different calls: a 'coo' and a 'grunt'. These calls are produced during affiliative social interactions. The coo call is a long tonal signal and is accompanied by a small mouth opening and protuding lips, whereas the grunt call is a noisy, pulsatile signal of short duration and is characterized by mouth opening with little or no lip protrusion. The coo sound was presented either in the context of the matching (congruent) visual recording of the monkey articulating a coo or in the context of a mismatching (incongruent) visual recording of the same monkey articulating a grunt. Note that the onset of the vocalization relative to the facial motion was kept identical across congruent and incongruent stimuli. The human stimuli were generated in a similar procedure and were as closely matched to the monkey stimulus material as possible. Specifically, the human actor was presented with the recordings of the monkey vocalizations and then instructed to produce sounds that were similar (including their duration). The incongruent human stimuli

were generated in a similar way to the monkey stimuli. All audio-visual monkey and human stimuli had a total duration of 3 seconds. This was achieved by adding a still frame taken from the start of the video and presenting it for 1 second before the onset of the vocalization and adding another still frame (last frame of the video recording) at the end of the stimulus and presenting it until 3 seconds were reached. By editing the material in this way, trials of equal duration were ensured.

Procedure

The infants were seated on their parent's lap in a dimly lit, sound-attenuated, and electrically shielded room. Prior to each presentation of a stimulus, an alerting sound was played in order to direct the infants' attention to the screen and then a fixation cross appeared in the center of the screen for 300 ms. All stimuli were projected at the center of the screen on a black background, using a 70 Hz, 17-inch computer screen at a distance of 60 cm from the eyes. The sound was presented via loudspeaker (SPL = 72 db for coos; 77 dB for grunts). All infants were presented with all four conditions (human-congruent, human-incongruent, monkey-congruent, monkey-incongruent). The stimuli were presented in a pseudo-random order with the constraint that none of the conditions was presented more than twice in a row. The inter-stimulus interval varied randomly between approximately 1000 and 2000 ms and during that time an abstract screensaver with moving shapes was presented on the screen in order to keep the infants' attention. Mothers were instructed to look down at the infant rather than at the computer screen. The session continued until the infant saw the maximum number of trials (200) or became fussy.

EEG measurement and data analysis

The EEG was recorded with Ag-AgCl electrodes from 19 scalp locations of the 10–20 system, referenced to Cz. Horizontal and vertical electro-oculograms (EOGs) were recorded bipolarly. Sampling rate was 250 Hz. EEG data were re-referenced to the algebraic mean of the left and the right mastoid electrodes, and band-pass filtered between 0.3 and 20 Hz (1501 points).

We carried out two analyses of the ERP data: (1) an analysis time-locked to the voice onset in order to assess multisensory (audio-visual) processing, and (2) an analysis time-locked to the face-onset (1000 ms prior to voice onset) in order to assess visual-specific processing of monkey and human faces. Data were baseline corrected by subtracting the average voltage in the 200 ms baseline period (i.e. 200 ms prior to face or voice onset) from each post-stimulus data point. For elimination of artifacts caused by eye and body movements, EEG data for the whole trial were rejected off-line whenever the standard deviation within a 200-ms gliding window exceeded 80 μV for the vertical or horizontal EOG and 50 μV at any other electrode. For statistical analysis of amplitude effects, mean amplitude was computed for 100 ms long time windows from 100 to 500 ms (zero corresponds to the vocalization onset). ERPs were evaluated by computing the following regions of interest (ROIs): fronto-central (Fz, Cz), temporal (T7, T8), occipital (O1, O2). These ROIs were chosen in order to assess effects related to frontal, auditory (temporal), and visual (occipital) information processing. The choice of ROIs and time windows was informed by prior infant ERP studies that have investigated multisensory perception (Hyde *et al.*, 2011; Kushnerenko, Teinonen, Volein & Csibra, 2008) and visual inspection of the data. It is important to note that while the selection of the ROIs was done in order to best represent information processes in auditory and visual cortices, the nature of the ERP signals is such that, due to volume conduction, the exact cortical generators might not be located in the cortical region above which the electrode is placed (DeBoer *et al.*, 2005). Because the morphology and timing of the ERP waveforms changes considerably during the first years of life, it is recommended that measurements should not be combined over more than 1-month intervals when conducting infant ERP studies (DeBoer *et al.*, 2005). Corroborating this idea, we obtained a main effect of age at occipital electrodes in a repeated measures ANOVA that assessed age (4 versus 8 months) as a between-subjects factor (100–200 ms: $F[1, 30] = 13.34$, $p = .001$, partial $\eta^2 = 0.31$; 200–300 ms: $F[1, 30] = 9.62$, $p = .004$, partial $\eta^2 = 0.24$). We therefore analyzed the data of both age groups separately. For the predefined ROIs, the ERP data were analyzed by repeated measures ANOVAs with the within-subject factors of (1) species (human versus monkey) and (2) congruence (congruent versus incongruent).

As mentioned above, we performed an analysis to examine potential differences in the visual processing of human and monkey faces by time-locking the ERP analysis to the onset of the face stimulus, which preceded the voice onset by 1000 ms. All other parameters of the ERP analysis were kept identical to the analysis that was time-locked to the voice onset. For statistical analysis of amplitude effects, mean amplitude was computed for two time windows that best captured the face-sensitive infant ERP components at occipital electrodes (O1 and O2): N290 (4-month-olds: 300–400 ms; 8-month-olds: 250–350 ms) and P400 (4-month-olds: 400–600 ms; 8-month-olds: 350–550 ms). For the occipital electrodes, the ERP data were analyzed by repeated measures ANOVAs with within-subject factors of (1) species (human versus monkey), (2) congruence (congruent versus incongruent), and (3) electrode (O1: left hemisphere versus O2: right hemisphere).

Results

Audio-visual processing of human and monkey stimuli

For our ERP analysis for the group of 4-month-old infants, an ANOVA with the factors species (human versus

monkey) and congruence (congruent versus incongruent) revealed a significant main effect for the factor congruence at temporal (auditory) electrodes between 200 and 300 ms (F[1, 15] = 6.61, p = .021, partial η^2 = 0.31) and occipital (visual) electrodes (F[1, 15] = 9.23, p = .008, partial η^2 = 0.38) (Figures 2 and 3). ERPs were more positive-going in the congruent condition than in the incongruent condition. The main effect of congruence continued to be significant at occipital electrodes from 300 to 500 ms: (300–400 ms: F[1, 15] = 10.46, p = .006, partial η^2 = 0.41; 400–500 ms: F[1, 15] = 13.33, p = .002, partial η^2 = 0.47). There was neither a main effect for the factor species (F[1, 15] = 0.53, p = .478, partial η^2 = 0.034 at temporal electrodes; F[1, 15] = 0.08, p = .78, partial η^2 = 0.034 at occipital electrodes) nor an interaction between the factors species and congruence (F[1, 15] = 0.17, p = .687, partial η^2 = 0.011 at temporal electrodes; F[1, 15] = 0.37, p = .55, partial η^2 = 0.024 at occipital electrodes). Thus, ERP responses from 4-month-old infants were sensitive to congruence but were insensitive to the different species. This is consistent with their multisensory matching behavior: 4-month-old infants match monkey faces to voices (Lewkowicz & Ghazanfar, 2006).

Our ERP analysis for the group of 8-month-old infants revealed a pattern of results that was different from the 4-month-olds. Eight-month-olds showed a significant interaction between the factors species (human versus monkey) and congruence (congruent versus incongruent) at occipital (visual) electrodes between 100 and 200 ms (F[1, 15] = 4.67, p = .047, partial η^2 = 0.24) and at fronto-central electrodes (F[1, 15] = 5.09, p = .039, partial η^2 = 0.25) (Figures 2 and 3). Specifically, only in the context of human faces and voices did 8-month-olds demonstrate significant differences between the congruent and incongruent condition (paired-sample comparisons between incongruent and congruent conditions in the human context at fronto-central electrodes: t[15] = 2.19, p = .044 and at occipital electrodes: t[15] = 2.33, p = .034). In the context of monkey faces and voices there was no difference between congruent and incongruent conditions (at fronto-central electrodes: t[15] = −1.14, p = .272 and at occipital electrodes: t[15] = −0.96, p = .35). That is, 8-month-old infant ERP responses were sensitive to the face–voice congruence of human signals that mimicked monkey vocalizations, but they were insensitive to congruence for monkey faces and voices. These data are consistent with behavioral perceptual narrowing data suggesting that 8-month-old infants can no longer match monkey faces and voices (Lewkowicz & Ghazanfar, 2006). The finding that the crucial effect was observed between 100 and 200 ms in 8-month-olds, whereas it was found between 200 and 300 ms in 4-month-olds is in line with prior perceptual work showing decreased ERP latencies as the brain matures in infancy (Csibra, Kushnerenko & Grossmann, 2008; Halit *et al.*, 2003; Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen & Näätänen, 2002). There was

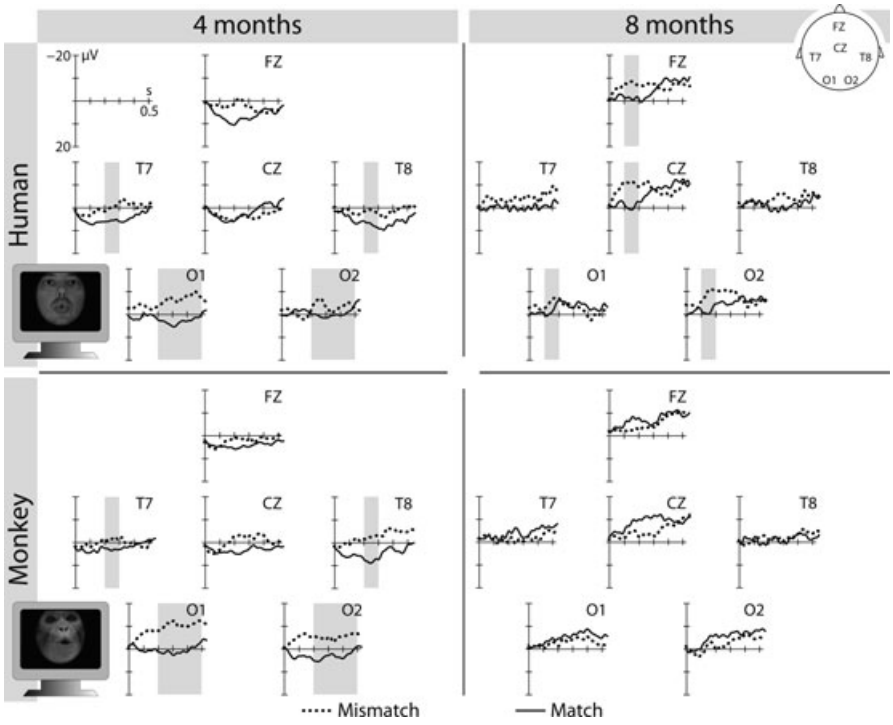

Figure 2 *This figure shows the event-related brain potentials (ERPs) time-locked to the voice onset (0 ms) in 4- and 8-month-old infants elicited by matching/congruent (solid line) and mismatching/incongruent (dotted line) human and monkey audio-visual stimuli. The time windows during which significant differences between the matching and mismatching condition were observed are marked in grey.*

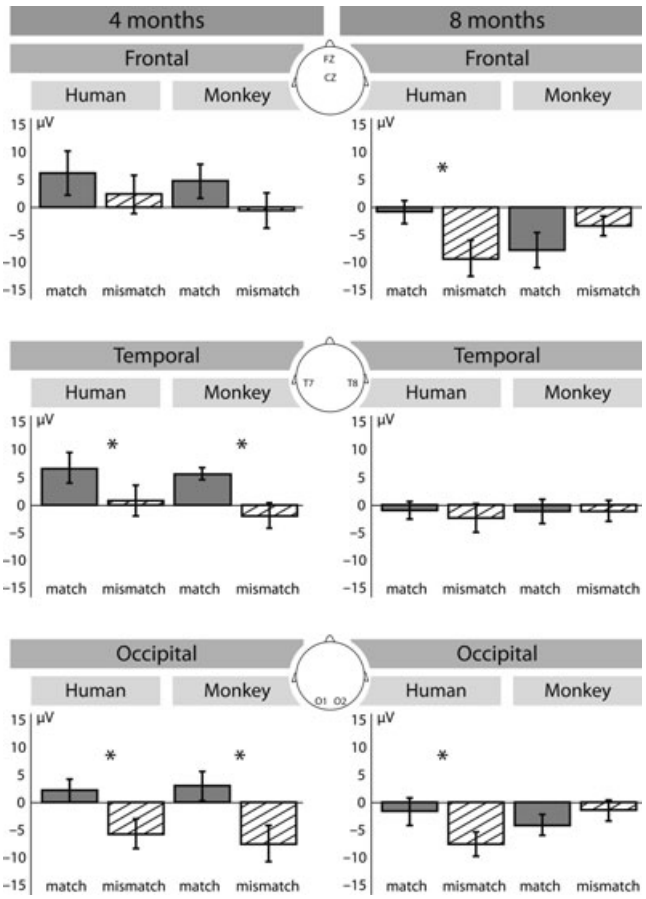

Figure 3 *This figure shows the mean amplitudes (±SE) of the ERPs (shown in Figure 2) elicited by matching and mismatching human and monkey audio-visual stimuli in 4- and 8-month-old infants (time window for which mean amplitude was computed: 4 months [200–300 ms]; 8 months[100–200 ms]). The significant differences between the matching and mismatching condition are marked by asterisks.*

neither a main effect of the factor species (F[1, 15] = 0.93, p = .35, partial η^2 = 0.058 at occipital electrodes; F[1, 15] = 0.078, p = .783, partial η^2 = 0.005 at fronto-central electrodes) nor a main effect of the factor congruence (F[1, 15] = 0.579, p = .459, partial η^2 = 0.037 at occipital electrodes; F[1, 15] = 0.393, p = .54, partial η^2 = 0.023 at fronto-central electrodes).

To test whether a long-range functional association between anterior and posterior brain sites for the processing of congruent multisensory information is established (constructed) with age (experience), we performed a correlation analysis between ERP responses from the fronto-central and occipital electrodes (mean amplitude during 200–300 ms in 4-month-olds and during 100–200 ms in 8-month-olds). Only for the 8-month-olds was there a positive correlation between ERP responses (mean amplitude between 100 and 200 ms) from the two sites (r = .487; p = .028). This was only observed for the human congruent condition. A similar association did not exist for any of the monkey conditions or for the human incongruent condition (all ps > .5). This suggests that multisensory perceptual narrowing may be the result of constructive (as opposed to regressive) neurodevelopmental events between frontal and occipital cortices present by the age of 8 months.

Visual processing of human and monkey faces

This ERP analysis revealed a main effect of the factor species on the N290 in 8-month-old infants (F[1, 15] = 5.58, p = .032, partial η^2 = 0.27): with monkey faces (M = 3.88 μV; SE = 3.26) eliciting a more negative-going N290 than human faces (M = 7.85 μV; SE = 3.04). In contrast, 4-month-old infants showed a more negative-going N290 to human faces (M = −8.89 μV; SE = 1.98) when compared to monkey faces (M = −6.35 μV; SE = 3.1) (Figure 4). However, ERPs elicited by human and monkey faces were not statistically different from each other (F[1, 15] = 1.097, p = .311, partial η^2 = 0.068). No effects were observed for the P400 component at either age. Moreover, there were no differences between the ERPs in response to monkey and human faces at the end of the face trial (800–1000 ms), which served as the baseline period for the ERP analysis locked to the voice-onset, at either age.

Discussion

We provide neurophysiological evidence for perceptual narrowing of face–voice matching during the first year of life. We measured ERPs while 4- and 8-month-old infants watched and listened to congruent and incongruent audio-visual presentations of monkey vocalizations and humans mimicking monkey sounds. Even though there has been some infant ERP work examining audio-visual processing of speech in 5-month-old infants (Hyde *et al.*, 2011; Kushnerenko *et al.*, 2008), ours is the first ERP study to use within- and across-species stimuli to test infants of different ages. We found that in younger infants, the ERP responses discriminated between the congruent and the incongruent faces and voices regardless of species, but in older infants, the ERP sensitivity to multisensory congruence was limited to the human face and voice. ERPs of 4-month-old infants measured over temporal (auditory) and occipital (visual) brain regions differed between congruent and incongruent trials across species, whereas 8-month-old infants' brain responses only differed over occipital regions in the human context. This demonstrates that although infants at both ages are sensitive to information across face and voice, the older infants' detection of audio-visual mismatches is specific to human stimuli, which is in line with the multisensory perceptual narrowing account (Lewkowicz & Ghazanfar, 2006, 2009).

The observed narrowing of the ERP responses measured over visual cortex is consistent with prior ERP work on the visual processing of faces during infancy. Similar to the current findings, infants' N290 component

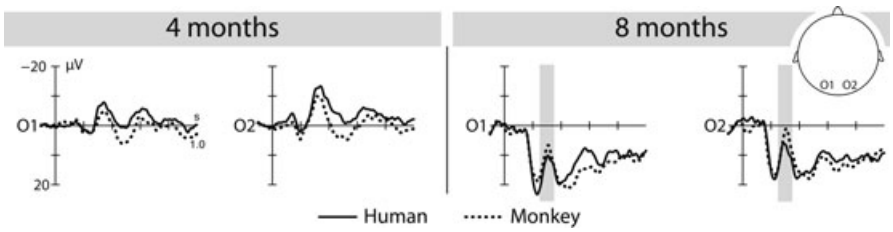

Figure 4 *This figure shows the ERPs time-locked to the face onset (0 ms) in response to human (solid line) and monkey (dotted line) faces for 4-month-old infants (top panel) and 8-month-old infants (bottom panel). The time window (face-sensitive infant N290: 250 to 350 ms) during which significant ERP differences between species were observed in 8-month-olds (marked by a rectangle) but not in 4-month-olds.*

(de Haan *et al*., 2002; Halit *et al*., 2004; Halit *et al*., 2003), which is considered to be the developmental precursor to the face-sensitive N170 in adults (Bentin *et al*., 1996), becomes tuned to human faces during the first year of life. Specifically, the infant N290 at 12 months, like the N170 in adults, is enhanced to inverted human but not to inverted monkey faces, whereas younger infants (3 and 6 months of age) showed no such effect (de Haan *et al*., 2002; Halit *et al*., 2004; Halit *et al*., 2003). Despite these similarities, the exact developmental relation between prior work on face processing and the ERP effects in the current study cannot easily be determined, since prior ERP work employed static pictures of faces, whereas in the current study dynamic audio-visual stimuli were used. However, a study conducted with 5-month-old infants that used matching (congruent) and mismatching (incongruent) dynamic audio-visual speech stimuli (Kushnerenko *et al*., 2008) reported a very similar timing to the ERP effects observed in our 4-month-old infants, suggesting similar temporal dynamics in the audio-visual integration process in young infants. Critically, prior infant ERP work used audio-visual speech stimuli that were highly familiar even to very young infants (Hyde *et al*., 2011; Kushnerenko *et al*., 2008), whereas we used humans that mimicked monkey grunt and coo vocalizations that were unfamiliar to the infants. Therefore, our findings suggest that young infants are sensitive to face-voice congruence even in fairly unfamiliar contexts. Moreover, using humans mimicking monkey sounds might be considered a stronger test of perceptual narrowing than using familiar human speech and comparing it to monkey vocalizations: Speech includes richer spectro-temporal cues (e.g. consonant-vowels, format shifts, etc.) that are absent in monkey calls (Rowell & Hinde, 1962).

In addition to the observed changes in auditory and visual face–voice processing, we found that 8-month-olds but not 4-month-olds showed a difference between the fronto-central ERPs elicited during congruent and incongruent human faces and voices. This suggests that, apart from sensory processes, older infants engage frontal brain areas when perceiving congruent versus incongruent human (but not monkey) face–voice pairings. This additional involvement of frontal resources at 8 months of age might reflect the emergence of more mature processing of face–voice correspondences. For example, young infants may match faces and voices based on simple intersensory cues by using mechanisms that are not specific to sensory signals in the social domain (Kuhl & Meltzoff, 1982; Patterson & Werker, 2003; Walton & Bower, 1984). As infants get older, however, and as they acquire greater experience with conspecific (i.e. same species) faces and voices, they become increasingly better at perceiving higher-level cues inherent in audio-visual facial expressions such as affect and gender (Flom & Bahrick, 2007; Grossmann, Striano & Friederici, 2006; Kahana-Kalman & Walker-Andrews, 2001; Patterson & Werker, 2002). In contrast to our findings, prior infant ERP work has reported frontal effects during multisensory processing at younger ages (5 months) (Hyde *et al*., 2011; Kushnerenko *et al*., 2008). However, as argued above, in these studies, infants were presented with highly familiar speech stimuli. Therefore, this discrepancy between studies may either be explained by increased familiarity with the stimulus material or by the fact that human speech engages specific frontal processes in early infancy (Dehaene-Lambertz, Hertz-Pannier, Dubois, Meriaux, Roche, Sigman & Dehaene, 2006).

For the congruent human condition, but not for any of the other conditions, we found a correlation between the brain responses recorded from fronto-central and occipital electrode sites in 8-month-old infants. This long-range functional correlation between frontal and occipital brain responses during congruent human trials might reflect information integration across frontal and occipital cortices. It further suggests that perceptual narrowing might be the result of constructive neurodevelopmental events across brain regions (Ghazanfar, 2012; Purves, White & Riddle, 1996; Quartz & Sejnowski, 1997) – in this case, the development of new functional connections between occipital and frontal regions of the brain (Gao, Zhu, Giovanello, Shen, Smith, Shen, Gilmore & Lin, 2009; Homae, Watanabe, Otobe, Nakano, Go, Konishi & Taga, 2010). This is in contrast with recent views suggesting that unisensory perceptual narrowing is due to the pruning of exuberant, unneeded synaptic connections (Scott, Pascalis & Nelson, 2007; but see Scott & Monesson, 2010).

It may be that both constructive and regressive neural developmental processes are at play during perceptual narrowing. It is difficult to assess either hypothesis with ERP data alone. However, in the present study, in addition to the functional connection between the frontal and the visual cortex that develops by 8 months of age, the timing of our ERP congruence effects also support a 'neuro-constructivist' account. Namely, younger infants showed an effect of congruence at temporal and occipital electrodes starting at around 200 ms, whereas the human-specific congruence effect in the older infants emerged earlier, after 100 ms at both occipital and frontal electrodes. Such a shortening in the latency of an ERP component during infancy has been observed in a number of auditory and visual ERP studies (see Csibra *et al.*, 2008). At the occipital electrodes only, the duration of the effect also decreased form 4 months to 8 months of age. Both neurophysiological changes could be due to increased myelination of fiber tracts (Barkovich, Kjos, Jackson & Norman, 1988; Csibra *et al.*, 2008) and/or an experience-dependent increase in the number of synapses (Ghazanfar, 2012; Purves *et al.*, 1996; Quartz & Sejnowski, 1997). Such neuroanatomical changes would result in decreased neural activity thresholds, leading to both faster and stronger neural responses. Furthermore, a constructivist interpretation of our data is consistent with a recent unisensory study of face processing in infants. When young human infants are trained by their caregivers to individuate monkey faces (by giving each monkey face a proper name), they do not undergo the typical perceptual narrowing of cross-species face processing (Scott & Monesson, 2009), and older infants develop face-specific ERP response to monkey faces that is lacking in younger infants; i.e. reflecting a neural specialization that is constructed (Scott & Monesson, 2010).

This raises the question of what kind of experience drives the perceptual narrowing effects in infants' ERPs. The current experimental design favors increased familiarity/experience with the visual stimuli (human face versus monkey face), but not the auditory stimuli (since both monkey and human sounds were unfamiliar to the infant), as one of the crucial aspects contributing to the observed developmental narrowing pattern in the present study. For example, the finding that there was an effect at occipital (visual) but not at temporal (auditory) electrodes in 8-month-olds may indicate that the observed developmental narrowing is accounted for by changes in visual cortex. However, it is unlikely that the observed effect can be explained by visual perceptual narrowing alone. In particular, we obtained a significant interaction between congruence and species at occipital and frontal electrodes in 8-month-old infants, while according to a pure visual perceptual narrowing account one would predict a main effect of species on the visual (occipital) processing of the stimuli presented, which is not supported by our data. Moreover, the observed developmental pattern of ERP findings is compatible with the view that, as brain development proceeds, auditory influences on visual circuits are more susceptible to experience than visual influences on auditory circuits (Gottlieb, 1971; Turkewitz & Kenny, 1982).

Our analysis of the visual processing of human and monkey faces showed that the infant N290, which is thought to be a precursor to the face-specific adult N170, is modulated by the factor species at the age of 8 months but not at the age of 4 months. More specifically, monkey versus human faces elicited an N290 that was larger in its amplitude. This suggests that the N290 becomes sensitive to species-specific information between 4 and 8 months of age, and is thus in line with the notion that the N290 increases in its specificity to human faces during the first year of life (de Haan *et al.*, 2002; Halit *et al.*, 2003). These findings are also consistent with the multisensory effects and differences between ages observed in our study. Namely, these findings concerning the visual processing of faces support the idea that the current experimental design favors increased familiarity/experience with the visual stimuli (human face versus monkey face), not the auditory stimuli (since both monkey and human sounds were unfamiliar to the infant), as one of the crucial aspects contributing to the observed developmental multisensory narrowing pattern in the present study.

In contrast to our findings, prior work has shown that while the specificity of the infant N290 for upright human faces increases during the first year of life, the amplitude of the N290 differs between monkey and human faces across infancy (3, 6, and 12 months of age), with human faces eliciting a larger N290 than monkey faces irrespective of face orientation (de Haan *et al.*, 2002; Halit *et al.*, 2003). These differences between the current findings and prior work might be due to methodological differences across studies (de Haan *et al.*, 2002; Halit *et al.*, 2003). For example, while prior studies (de Haan *et al.*, 2002; Halit *et al.*, 2003) presented stimuli that included external features apart from the face, our stimuli were edited in such a way that external features of the head such as hair and ears were not visible. Presenting external features that critically differed between humans and monkeys may have helped infants in prior studies to discriminate between species, while the discrimination between human and monkey faces solely on facial features is only achieved once the infants' perceptual system has become more finely tuned to human faces.

Finally, we would like to consider a few remaining issues that arise from the current ERP findings that deserve attention and should be more closely examined in future studies. The first issue has to do with clarifying whether the ERP responses observed in the current study reflect multisensory *integration* or are merely the result of additive unisensory processes. One way to address this question is to examine whether there are super-additive ERP effects in which the multisensory response exceeds the sum of those evoked by the modality-specific (auditory and visual) stimulus components. Our face–voice

matching paradigm does not allow us to assess whether the ERP responses are super-additive in nature. To our knowledge, the only infant ERP study that attempted to address this question, while showing that visual stimulation enhances auditory processing as indexed by an enhanced N450 at fronto-temporal sites, failed to find super-additive effects in 3-month-old infants' ERPs (Hyde *et al*., 2010).

Another issue that needs consideration is how the current findings relate to a recent study by Vogel and colleagues (2012) that investigated how familiarity with race affects voice–face matching. The ERP results of this study revealed that while 5-month-old infants were able to match vocal emotional expressions (crying or laughing) to a subsequently presented facial expression (sad or happy face) for familiar race faces and unfamiliar race faces alike, 9-month-old infants' matching was limited to the familiar race faces. These findings are broadly consistent with our ERP findings and together these two studies provide converging evidence for perceptual narrowing in face–voice matching during infancy. However, there are some differences between the two studies, particularly with respect to the topography of the ERP effects observed across ages. Vogel and colleagues (2012) found a shift from anterior (fronto-central) to posterior (occipital) with increasing age of the infants, while the current data suggests a different pattern with an anterior component emerging with increasing age and experience. There are a number of methodological factors that might account for this difference across studies. For example, in the current study, we employed a face–voice matching paradigm and the facial information was available while infants processed the vocal information, whereas in Vogel *et al*.'s (2012) study the voice was presented first and then the facial expression was presented afterwards. While these methodological differences may play a role in accounting for these differences, there is also a conceptual difference between these studies. While Vogel *et al*. (2012) contrasted own-species familiar race face stimuli to unfamiliar race face stimuli, in the current study, stimuli crossed the species divide by contrasting humans to non-human animals (monkeys). It seems likely that the observed differences across studies might be explained by the fact that with increasing age and learning experience infants treat humans as a special category and differentiate this category from other non-human species possibly at a global level of exclusivity (see Quinn, 2004). Thus, distinguishing between different race stimuli and matching information across face and voice within the familiar human category may rely more on fine perceptual distinctions as seen in a shift to more posterior processing in older infants in Vogel *et al*.'s (2012) study, while distinguishing between different species stimuli may rely more on an increasing knowledge of the own species as seen in the integration of anterior and posterior brain processes in the current study. Clearly, this is only a tentative proposal that should be explicitly tested in future work.

Taken together, the present data indicate that during the ages when multisensory perceptual narrowing occurs in human infants, their brains undergo a corresponding neurophysiological change. As infants get older, their visual and frontal brain regions (and their functional connection) become more sensitive to congruence of human faces and voices relative to monkey faces and voices – even though the humans were mimicking monkey calls. This suggests that experience with species identity (via indexical cues), independent of any semantic content, contributes to these neural changes. The neural changes seem to be the result of constructive neurodevelopmental events as opposed to regressive ones.

References

Bahrick, L.E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. *Advances in Child Development and Behavior*, **30**, 153–187.

Barkovich, A.J., Kjos, B.O., Jackson, D.E., & Norman, D. (1988). Normal maturation of the neonatal and infant brain: MR imaging at 1.5 T. *Radiology*, **166**, 173–180.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, **8**, 551–565.

Csibra, G., Kushnerenko, E., & Grossmann, T. (2008). Electrophysiological methods in studying infant cognitive development. In C.A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd edn., pp. 247–262). Cambridge, MA: MIT Press.

de Haan, M., Pascalis, O., & Johnson, M.H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, **14**, 199–209.

DeBoer, T., Scott, L.S., & Nelson, C.A. (2005). ERPs in developmental populations. In T.C. Handy (Ed.), *Event-related potentials* (pp. 263–297). Camrbidge, MA: MIT Press.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M., & Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 14240–14245.

Flom, R., & Bahrick, L.E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: the role of intersensory redundancy. *Developmental Psychology*, **43**, 238–252.

Gao, W., Zhu, H., Giovanello, K.S., Shen, D., Smith, J.K., Shen, D., Gilmore, J.H., & Lin, W. (2009). Evidence on the emergence of the brain's default network from 2-week-old to 2-year-old healthy pediatric subjects. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 6790–6795.

Ghazanfar, A.A. (2012). The evolution of multisensory vocal communication in primates and the influence of developmental timing. In A.J. Bremner, D.J. Lewkowicz & C. Spence (Eds.), *Multisensory development* (pp. 360–371). Oxford: Oxford University Press.

Gottlieb, G. (1971). Ontogenesis of sensory function in birds and mammals. In E. Tobach, L.R. Aronson & E. Shaw (Eds.), *The biopsychology of development* (pp. 67–128). New York: Academic Press.

Grossmann, T., Striano, T., & Friederici, A.D. (2006). Crossmodal integration of emotional information across face and voice in the infant brain. *Developmental Science*, **9**, 309–315.

Halit, H., Csibra, G., Volein, A., & Johnson, M.H. (2004). Face-sensitive cortical processing in early infancy. *Journal of Child Psychology & Psychiatry*, **45**, 1228–1234.

Halit, H., de Haan, M., & Johnson, M.H. (2003). Cortical specialization for face processing: face-sensitive event-related potential components in 3- and 12-month-old infants. *NeuroImage*, **19**, 1180–1193.

Homae, F., Watanabe, H., Otobe, T., Nakano, T., Go, T., Konishi, Y., & Taga, G. (2010). Development of global cortical networks in early infancy. *Journal of Neuroscience*, **53**, 4877–4882.

Hyde, D.C., Jones, B.L., Flom, R., & Porter, C.L. (2011). Neural signatures of face–voice synchrony in 5-month-old human infants. *Developmental Psychobiology*, **53**, 359–370.

Hyde, D.C., Jones, B.L., Porter, C.L., & Flom, R. (2010). Visual stimulation enhances auditory processing in 3-month-old infants and adults. *Developmental Psychobiology*, **52**, 181–189.

Kahana-Kalman, R., & Walker-Andrews, A.S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development*, **72**, 352–369.

Kuhl, P.K., & Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. *Science*, **218**, 1138–1141.

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002). Maturation of auditory event-related potentials during the first year of life. *NeuroReport*, **13**, 47–51.

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 11442–11445.

Lewkowicz, D.J., & Ghazanfar, A.A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 6771–6774.

Lewkowicz, D.J., & Ghazanfar, A.A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, **13**, 470–478.

Pascalis, O., de Haan, M.H., & Nelson, C.A. (2002). Is face processing species-specific during the first year of life? *Science*, **296**, 1321–1323.

Patterson, M.L., & Werker, J.F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, **81**, 93–115.

Patterson, M.L., & Werker, J.F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, **6**, 191–196.

Purves, D., White, L.E., & Riddle, D.R. (1996). Is neural development Darwinian? *Trends in Neurosciences*, **19**, 460–464.

Quartz, S.R., & Sejnowski, T.J. (1997). The neural basis of cognitive development: a constructivist manifesto. *Behavioral and Brain Sciences*, **20**, 537–556.

Quinn, P. (2004). Is asymmetry in young infants' categorization of humans versus nonhuman animals based on head, body, or global gestalt information? *Psychonomic Bulletin & Review*, **11**, 92–97.

Rowell, T.E., & Hinde, R.E. (1962). Vocal communication by the rhesus monkey (*Macaca mulatta*). *Proceedings of the Zoological Society London*, **138**, 279–294.

Scott, L.S., & Monesson, A. (2009). The origin of biases in face perception. *Psychological Science*, **20**, 676–680.

Scott, L.S., & Monesson, A. (2010). Experience-dependent neural specialization during infancy. *Neuropsychologia*, **48**, 1857–1861.

Scott, L.S., Pascalis, O., & Nelson, C.A. (2007). A domain general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, **16**, 197–201.

Turkewitz, G., & Kenny, P.A. (1982). Limitations on input as a basis for neural organization and perceptual development: a preliminary theoretical statement. *Developmental Psychobiology*, **15**, 357–368.

Vogel, M., Monesson, A., & Scott, L.S. (2012). Building biases in infancy: the influence of race on face and voice emotion matching. *Developmental Science*, **15**, 359–372.

Wallace, M.T., Perrault, T.J., Hairston, W.D., & Stein, B.E. (2004). Visual experience is necessary for the development of multisensory integration. *Journal of Neuroscience*, **24**, 9580–9584.

Wallace, M.T., & Stein, B.E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience*, **17**, 2429–2444.

Wallace, M.T., & Stein, B.E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *Journal of Neuroscience*, **21**, 8886–8894.

Wallace, M.T., Stein, B.E., & Ramachandran, R. (2006). Early experience determines how the senses will interact: a revised view of sensory cortical parcellation. *Journal of Neurophysiology*, **101**, 2167–2172.

Walton, G.E., & Bower, T.G. (1984). Amodal representations of speech in infants. *Infant Behavior and Development*, **7**, 49–63.

Zangenehpour, S., Ghazanfar, A.A., Lewkowicz, D.J., & Zatorre, R.J. (2009). Heterochrony and cross-species intersensory matching by infant vervet monkeys. *PLoS ONE*, **4**, e4302.

Received: 23 November 2011
Accepted: 5 June 2012